

# Linguistic Representation of the *Sewol Ferry Disaster*

## : A Corpus-based Analysis of Headlines from Two Korean Newspapers<sup>1</sup>

You-kyung Ju (SOAS) & Jaehoon Yeon (SOAS)

### Abstract

This study is a corpus-based analysis of headlines about the *Sewol Ferry Disaster* from two Korean newspapers, *Chosun ilbo* and *Hankyoreh shinmun*. The disaster occurred in April 2014, and a special law for the *Sewol Ferry Disaster* was enacted in October 2014. The headlines of the two Korean newspapers from April 2014 to October 2014 were collected and examined for linguistic peculiarities in the word frequency in order to identify the ideological inclinations of the two newspapers. The corpus-based linguistic study revealed that the headlines have concealed ideological meanings and are manipulated along some ideological lines to use words that reflect newspaper companies' views and perceptions. The study concludes that headlines can be used as a perspective-inducing strategy by the editors of newspaper companies in order to create and foster the readers' opinions on national issues, and eventually shift the responsibility for the disaster in this case.

### 1. Introduction

Linguists' interest in discourse has gradually shifted from the traditional focus on the linguistic structure of texts to how texts figure in social interaction. An understanding of the phonology, morphology, grammar and semantics of a text does not necessarily constitute an understanding of the text. Instead, the rhetorical intention, the conception, the attitudes and values that the author and receptor bring to the text are equally essential (Kaplan, 1990). Language is thus no longer seen as merely reflecting our reality, but as central to interpreting or even creating reality. Beard (2000: 18) puts it like this:

Language is a means of communication, a means of presenting and shaping series of beliefs. Language is not something somehow separate from the ideas it contains, but the way language is used says a great deal about how the ideas have been shaped.

There are always different ways of describing the same thing and differences in expression may therefore carry ideological distinctions. In other words, when used efficiently, language is the most powerful manipulative tool that influences the public.

---

<sup>1</sup> This work was supported by the Academy of Korean Studies Grant (KSPS) funded by the Korean Government (MOE) (AKS-2011-BAA-2014).

The two domains in society most accused of using language manipulatively are probably politics and journalism. Opinion leaders, governments, newspaper editors, etc., play a crucial role in shaping issues in society and setting the boundaries of what is talked about and how it is talked about (Henry and Tator, 2002). Newspapers are particularly known to lead in the initiation of discourse on key national issues, such as disasters, by controlling the content and form of information transmitted and exposing them to their readership. The initiation, advancement, and sustenance of discourse by newspapers often contain some embellishments, determined by their own angle of the story. Journalists working at newspaper companies therefore act as gate keepers, interpreters, and commentators on the issues, not mere passive “information channels” (Anzur, 2000).

This paper is an attempt to look at how language is used in Korean newspapers, particularly in headlines, to reflect specific ideologies and power relations with regard to the disaster management of the *Sewol Ferry Disaster* which occurred on 16 April 2014. The role headlines play as precursors to the news causes them to be constructed in such a way that they usually evoke readers’ emotions. Most editors ensure that the choice of words and expressions in headlines reflects the feelings, opinions and attitudes of people about issues in the news. To investigate such emotion-inducing strategies appearing in headlines and to identify the ideological inclinations that lie behind their constructions, a large corpus of headline data relating to the *Sewol Ferry Disaster* (hereafter, *SFD*) from two Korean newspapers, *Chosun ilbo* and *Hankyoreh shinmun*, was collected and analyzed by employing a corpus-based approach. By investigating the frequency of associated words of the search term *Sewol Ferry*, we provide evidence of salient patterns of linguistic peculiarity surrounding the disaster in two different newspapers with different ideological inclinations.

## **2. A corpus-based analysis of news discourse**

### **2.1. Corpus Linguistics**

In the current study, we intend to address how language may reflect underlying ideologies by examining the frequency of the related words to the *SFD* in large corpora composed of data from two Korean newspapers, by employing the methods and tools of Corpus Linguistics.

Corpus Linguistics is a quantitative method of carrying out linguistic analyses and has become one of the most wide-spread methods of linguistic investigation in recent years (Baker et al., 2008). Particularly, Corpus Linguistics has been increasingly used in critical discourse studies to examine the discourse of racism (Krishnamurthy, 1996), homosexuality (Baker, 2004), climate change (Grundmann and Krishnamurthy, 2010), refugees (Baker and McEnery, 1996; Gabrielatos and Baker, 2008) and gender issues (Hunt, 2011), uncovering the existence of language patterns and the

messages in them. By doing so, Corpus Linguistics shows how such empirical investigations of corpora can shed new light on previously intractable questions in linguistics.

A corpus is primarily a systematic collection of naturally occurring texts of both written and spoken language. It is often restricted to certain text types, to one or several varieties of the language, and to a certain time span. In a narrower sense nowadays, the term corpus is often only used to refer to systematic text collections that have been computerized. The analysis is therefore usually performed with the help of the computer, i.e. with specialised software, and takes into account the frequency of the phenomena investigated.

As software tools in corpus analysis process several batches of text at a time and rapidly present manually unmanageable numbers of occurrences of a given object, a corpus-based approach to discourse analysis is illuminating for describing collocational and other recurrent patterns associated with specific lexical items across an entire corpus. A corpus-based approach is indeed instrumental in gaining considerable degree of consensus about what is found in the examination of texts with statistical findings drawn from a larger corpus.

## **2.2. Language as the representative of ideological systems**

A corpus-based approach to discourse analysis basically presumes that language is not just a transparent medium of communication about the objective world, but a constantly operating representative of the prevailing ideological systems in the society. Analyzing discourse, therefore, eventually aims to find the hidden ideological meanings of the systematic social representation.

Language users tend to prefer certain lexical choices over others when referring to a certain phenomenon or a group of people and, in so doing, will propagate a certain version of reality, although they normally have a range of lexical items at their disposal (Jaworska & Krishnamurthy, 2012). Van Dijk (1995) highlights the fact that lexical choices are not only evaluative judgements, but also reflect the ideological stance of the user, or the group that she or he represents. Their choice of words, then, may evoke certain feelings in their readers or audiences and, in that way, direct their opinions.

In news texts, ideology is expressed in many disciplines with different, but overlapping shares of meaning by journalists who prefer particular lexical choices over others. In this respect, Fowler claims that news is not a value-free reflection of 'facts' and the structure of a news text embodies values and beliefs (Fowler, 1991). In the same light, Beard argues that there is no such thing as an unbiased report, and no such thing as 'neutral' language (Beard, 2000:18). The frequent patterns of the lexical choices in news texts are therefore especially interesting to investigate as they can point to the salient ideas, concepts and beliefs associated with a particular phenomenon in a society. In doing so, they can provide indications as to how the phenomenon is frequently framed by journalists in news discourse.

By investigating the collocation profiles of the search term *Sewol Ferry* in two large corpora of newspaper headlines, we will provide attested evidence of salient discourse patterns, and consequently the discourse prosodies surrounding the word *Sewol Ferry*, in two different ideological contexts.

### 2.3. Newspaper headlines

The power that newspaper headlines wield as the most significant news message comes naturally with discourse (MacRitchie & Seedat, 2008) - newspaper headlines act as forerunners to news reports and enjoy a privileged position that is explicitly reflected by a distinct layout and typography (White, 2011). In other words, the power of headlines is that they are the compact representation of “a set of meanings, metaphors, representations, images, stories, and statements that produce a particular version of events” (Burr, 1995: 339). Particularly, as headlines reveal the social, cultural and national representations circulating within a society at any given time, they reach an audience considerably wider than simply those who actually read the news story.

According to Taiwo (2004: 324), headlines are strategically used by editors, who choose emotive vocabulary, and rhetorical and graphological devices to make an impact on the readership. Newspaper headlines therefore present what newspaper companies regard as the most significant information readers ought to know (Metila, 2013). In news discourse, newspaper companies can limit or enable the construction of ‘reality’ in a certain way through a deliberate choice of headline information to present a narrow or one-sided view (Louw, 2001). This power to censor or filter information demonstrates the power of newspaper companies to preserve or abolish the hegemony of social groups (Metila, 2013).

### 3. Research questions and methodology

To explore how newspaper headlines represent what newspaper companies regard as most crucial regarding the *SFD*, we collected headlines from the websites of two newspaper companies, *Chosun ilbo* and *Hankyoreh shinmun*. *Chosun ilbo* is well-known as one of the most conservative newspapers in Korea and *Hankyoreh shinmun* for its liberal inclination.

The research questions of the present corpus-based analysis of newspaper headlines have arisen from the assumption that the frequency of a certain word (whether it be a person, governmental organization or a private company) would be associated with who the editors of these newspapers are presenting as responsible the *SFD*. Based on this assumption, we have examined the following questions, which are schematized in Figure 1:

- (1) Does the ideology of the newspaper company affect word frequency?
- (2) Does the frequency of a certain word or a group of words the newspaper company particularly choose to use differ depending on the words' relevance to government, i.e. whether it is government-related or non-government-related?
- (3) Finally, is the word frequency between government-related and non-government-related words, which we assume is associated with the responsibility for the *SFD*, different depending on the newspaper company's conservative or liberal inclination?

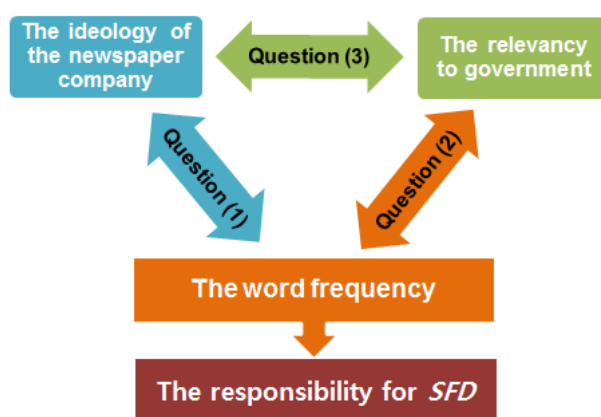


Figure 1. The correlations of the research questions

To investigate linguistic representations of the *SFD* in media discourse, we have conducted a corpus-based study of newspaper headlines. Two corpora of newspaper headlines that appeared in *Chosun ilbo* and *Hankyoreh shinmun* between 16 April 2014, the day when the *SFD* happened, and 16 October 2014, six months after the *SFD*, were compiled. The six-month time span of data collection corresponds to the disaster period (the first three months) and the special law period (the subsequent three months), due to the significant amount of data produced in the initial six months following the disaster.

Over this six month time span, we retrieved the word ‘*Sewol Ferry*’ from the online websites of the two newspaper companies, *Chosun ilbo* ([www.chosun.com](http://www.chosun.com)) and *Hankyoreh shinmun* ([www.hani.co.kr](http://www.hani.co.kr)) and archived the corpus data of the news article headlines. The keyword ‘*Sewol Ferry*’ had to appear either in the headline or in the article. For data analysis, *QSR Nvivo*, one of the leading qualitative research tools, was used to investigate word frequency.

#### 4. Findings and discussion

The data we collected seemed very intriguing even from an initial stage of analysis. The total number of headlines relating to the term *Sewol Ferry* in the two newspapers is significantly different – while

the total number of headlines in *Chosun* was 2,024, we found 5,147 headlines in *Hankyoreh*, which is more than double the number of headlines found in *Chosun*. Regarding our research question (1), the higher number of headlines in relation to the *SFD* seems to indicate the greater weight, or attention, which the newspaper company puts on the issue.

As for the corpus-based analysis, we used *QSR Nvivo* to investigate the word frequency patterns in the headlines of the two different newspapers. The Word Frequency Query of *QSR Nvivo* shows the results as in Table 1.

Word (조선)	Length	Count	Percentage (%)	Word (한겨레)	Length	Count	Percentage (%)
세월호	3	483	1.87	세월호	3	931	1.43
참사	2	182	0.71	청와대	3	251	0.39
유병언	3	110	0.43	대통령	3	220	0.34
침몰	2	92	0.36	국회	2	204	0.31
박대통령	4	67	0.26	정당	2	175	0.27
수사	2	59	0.23	유병언	3	127	0.20
진도	2	55	0.21	새누리	3	122	0.19
여객선	3	50	0.19	새정치	3	92	0.14
홍리	2	37	0.14	참사	2	90	0.14
해경	2	37	0.14	종교	2	89	0.14
지방선거	4	36	0.14	해경	2	89	0.14
유족	2	35	0.14	유족	2	87	0.13
국민	2	32	0.12	특별법	3	87	0.13
구조	2	30	0.12	실종자	3	82	0.13
대통령	3	30	0.12	검찰	2	79	0.12
국회	2	29	0.11	단독	2	74	0.11
정부	2	29	0.11	정부	2	70	0.11
사고	2	26	0.10	유가족	3	69	0.11
선택	2	26	0.10	사고	2	65	0.10
합의	2	20	0.08	속보	2	62	0.10

Table 1. Word Frequency Query in search for ‘Sewol Ferry’ in *Chosun* and *Hankyoreh*

The four columns in the left-hand side of the table are from *Chosun* and the next four columns of the table are from *Hankyoreh*. ‘Length’ means the number of syllables we set for our search – we put 2 syllables as the minimum number of syllables. So we can see 2, 3, and 4 syllable-words in the second and sixth columns. ‘Count’ means the frequency of the word in the data set. The ‘Weighted Percentage (%)’ is the part we need to pay attention to – it is the proportion of the word’s frequency in the whole data set. For example, the first line of the *Chosun* section of the table shows that the word ‘세월호 (*Seywelho*; *Sewol Ferry*)’ appeared 483 times in the whole data set and its word frequency makes up 1.87 % of the whole data set.

Having excluded section titles and numbers marking dates and the pages of the newspaper articles, we have subsequently extracted the following 12 most common words in the headlines of the two newspapers, in addition to the main search word ‘세월호 (*Seywelho*; *Sewol Ferry*)’: 참사 (*chamsa*;

disaster), 유병언 (*Yu Byengeon*; the president of the Semo group, the parent company of the *Sewol* Ferry), 침몰 (*chimmol*; the sinking), 대통령 (*taythonglyeng*; the President), 해경 (*haykyeng*; National Coast Guard), 유족 (*yucok*; the bereaved family), 국회 (*kwukhoy*; The National Assembly), 정부 (*cengpwu*; the government), 사고 (*sako*; the accident), 교황 (*kyohwang*; the Pope), 유가족 (*yukacok*; the family of the deceased), 청와대 (*chengwatay*; the Blue House) and 후보 (*hwupo*; the candidates).

To investigate the extent to which government-related words and responsibility for the *SFD* are associated with their respective word frequencies regarding our research question (2), we divided the 10 most common words into two groups as shown in Figure 2<sup>2</sup>:

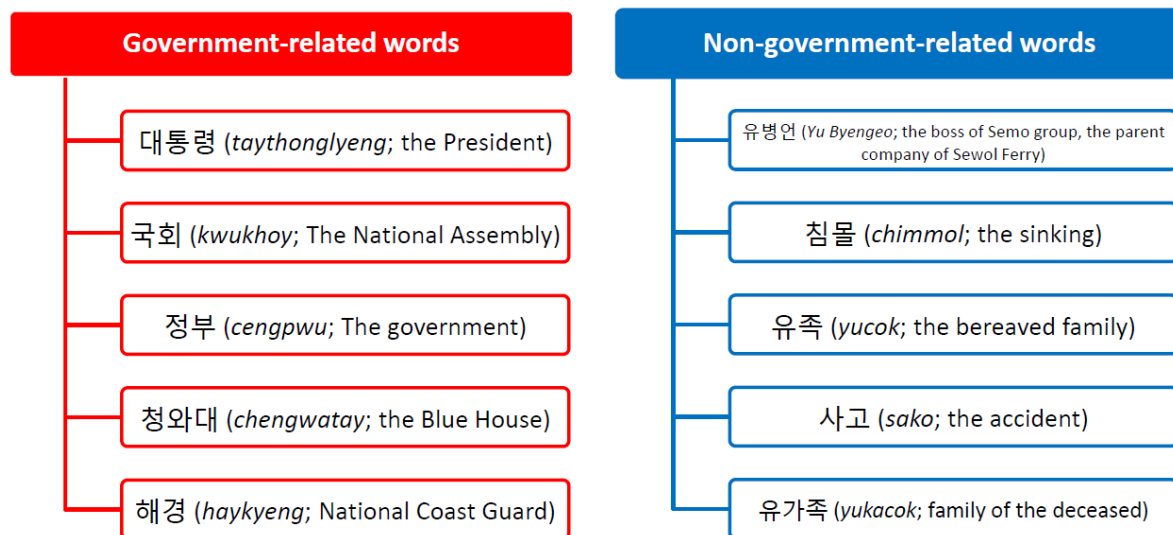


Figure 2. A pair of government-related and non-government-related words which commonly appeared in *Chosun* and *Hankyoreh* in relation to the search term *Sewol Ferry*

IBM SPSS (Statistical Package for the Social Science) Version 20 was used for analyzing the data. ‘Factorial ANOVA’ was conducted to explore our research question (3), which examines the impact of the ideology of newspaper companies and the influence of government factors on the weight of the most frequent words appearing in the headlines as measured by the Word Frequency Query of *QSR Nvivo*. The two categorical independent variables are therefore the ideology of the newspaper companies (conservative and liberal inclinations) and the influence of the governmental factor (government-related and non-government-related words), and the continuous dependent variable is the weighted percentage of the 10 common words appearing in the two newspapers as in Table 2:

<sup>2</sup> Among the 12 most common words (excluding the main search word 세월호 (*Seywelho*; *Sewol Ferry*)), 교황 (*kyohwang*; the Pope) and 후보 (*hwupo*; the candidates) are excluded from grouping because they are not relevant to the government relevancy factor.

**Between-Subjects Factors**

		Value Label	N
Ideology	1.00	Chosun	10
	2.00	Han	10
Governmental	1.00	Governmental	10
	2.00	Non-governmental	10

Table 2. The two categorical independent factors of the Factorial ANOVA

The statistical analysis indicates whether there is a change in weight of the most frequent words depending on the ideology of the newspaper companies (the main effect of ideology regarding our research question (1)). It also compares the weight of the most frequent words in two newspaper companies in terms of the influence of the governmental factor (the main effect of the governmental factor regarding our research question (2)). Finally, it shows whether the change in weight over the newspaper companies' ideologies is different for the government factors (the interaction effect regarding our research question (3)). The means of the weights of the frequent words of the two newspaper companies over government factors are presented in Table 3:

**Descriptive Statistics**

Dependent Variable: Weight

Ideology	Governmental	Mean	Std. Deviation	N
Chosun	Governmental	.1080	.02950	5
	Non-governmental	.2200	.16355	5
	Total	.1640	.12554	10
Han	Governmental	.2580	.12518	5
	Non-governmental	.1260	.04393	5
	Total	.1920	.11253	10
Total	Governmental	.1830	.11662	10
	Non-governmental	.1730	.12329	10
	Total	.1780	.11692	20

Table 3. Means of the weights of the frequent words of the two newspaper companies over government factors

All the effects were reported as significant at  $p < .05$ . There isn't a significant main effect of ideology on the weight of frequent words. This means that overall, when we ignore whether the word is government-related and non-government-related, the ideology of the newspaper company did not influence the weighted percentage of these words. Likewise, the main effect of the government



relevancy factor is not significant either, which means that, regardless of the newspaper company being conservative or liberal, government relevancy did not affect the weighted percentage of words either.

However, this output does suggest a significant result for the interaction between the effect of the ideology the newspaper company holds and the effect of the government relevancy of the word used in their newspaper articles in relation to the *Sewol Ferry*,  $F(1, 16) = 6.58, p < .05$ . This result means that the effect of government relevancy on the weighted percentage of the words was different in *Chosun* compared to *Hankyoreh*. The interaction effect is clearly shown in Figure 2:

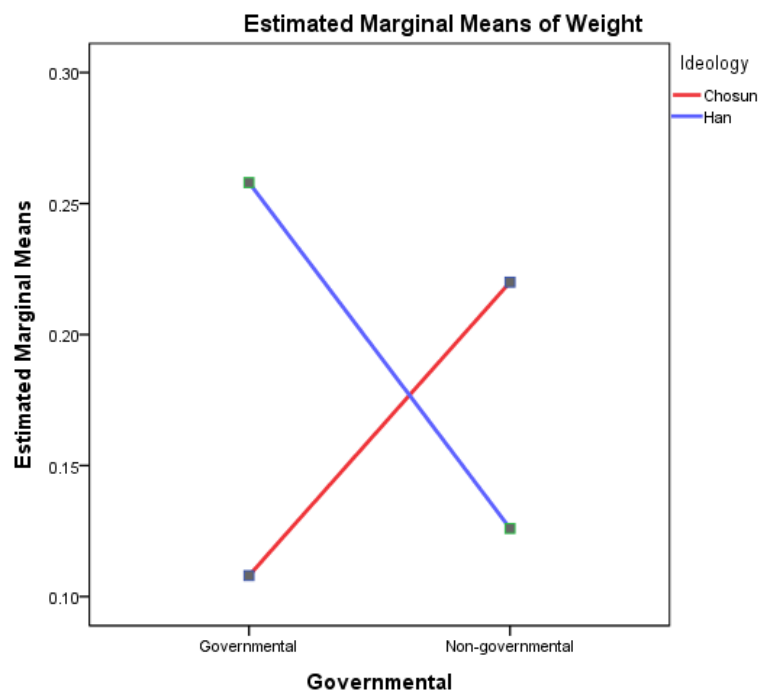


Figure 2. The interaction between the effect of the ideology the newspaper company holds and the effect of the government relevancy of the word used in their newspaper articles in relation to the *Sewol Ferry*

Non-parallel lines usually indicate a significant interaction effect and, in this particular graph, the lines actually cross, which indicates a large interaction between the independent variables – the government relevancy of the words and the ideology of the two newspaper companies. The results show that the weighted percentage of non-government-related words is higher in *Chosun*, whereas government-related words appeared more frequently than non-government-related words in *Hankyoreh*.

## 5. Conclusion

This study has observed how the ideologies of newspaper companies work in the construction of newspaper headlines. Working within the discipline of corpus analysis of newspaper discourse, our findings have demonstrated that there are certain hidden political inclinations behind the apparently simple statements of newspaper companies. An analysis of the search term *Sewol Ferry Disaster* from two Korean newspapers, *Chosun ilbo* and *Hankyoreh shinmun*, in a corpus revealed that there is a strong tendency to portray the extent of government relevancy to the *SFD* differently depending on the newspaper company's different ideological inclinations. In both corpora, there is a general association of ideology with governmental factors in the word frequency.

We return to our earlier hypothesis that word frequency is closely related to the agents' perception of the degree of responsibility the word bears for the *SFD*. One interesting point that this data demonstrates is that *Chosun* might have shifted responsibility for the *Sewol Ferry Disaster* to non-government-related people, organizations, or situations, such as 유병언 (*Yu Byengeon*; the president of the Semo group, the parent company of Sewol Ferry), 침몰 (*chimmol*; the sinking), 유족 (*yucok*; the bereaved family), 사고 (*sako*; the accident), and 유가족 (*yukacok*; the family of the deceased), whereas *Hankyoreh* has presented a situation in which the current *Park Kunhye* government, the ruling *Saenuri* party, the National Coast Guard, the Blue House, and the National Assembly did not react properly to the disaster and so bear more responsibility.

Therefore, we can conclude that a single news item, the *Sewol Ferry Disaster* in this study, is presented in different newspapers in different perspectives, keeping their political inclinations in view. The corpus-based linguistic study has rendered transparent the concealed ideological inclinations of the two newspapers, which were opaque for casual readers. The study concludes that headlines can be used as a perspective-inducing strategy by the editors of newspaper companies in order to create and foster readers' opinions on national issues, and eventually shift the responsibility for the disaster in this case.

## References

- Anzur, T. (2000). How to Talk to the Media: Televised Coverage of Public Health Issues in a Disaster. *Journal of Communication*, 15: 196-198.
- Baker, P. (2004). 'Unnatural Acts': Discourses of homosexuality within the House of Lords debates on gay male law reform. *Journal of Sociolinguistics*, 8(1): 88-106.
- Baker, P. & McEnery, T. (1996). A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Journal of Language and Politics*, 4(2): 197-226.
- Baker, P., Gabrielatos C., KhosraviNik, M., Krzyzanowski, M., McEnery, T. & Wodak, R. (2008) A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to

- examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3): 273-306.
- Beard, A. (2000). *The Language of Politics*. London: Routledge.
- Burr, V. (1995). *An Introduction to Social Construction*. London: Routledge.
- Fowler, R. (1991). Language in the News. *Discourse and Ideology in the Press*. London: Routledge.
- Gabrielatos, C. & Baker, P. (2008). Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK Press, 1996-2005. *Journal of English Linguistics*, 36: 5-38.
- Grundmann, R. & Krishnamurthy, R. (2010). The discourse of climate change: A corpus based approach. *Critical Approaches to Discourse Analysis Across Disciplines*, 4(2): 125-146.
- Henry, F. & Tator, C. (2002). *Discourses of Domination: Racial Bias in the Canadian English-Language Press*. University of Toronto Press.
- Hunt, S. (2011). *The discursial construction of female physical identity in selected works in children's literature*. Unpublished doctoral thesis, Rhodes University, Grahamstown, South Africa.
- Jaworska, S. & Krishnamurthy, R. (2012). On the F word: A corpus-based analysis of the media representation of feminism in British and German press discourse, 1990-2009, *Discourse & Society*, 23(4): 401-431.
- Kaplan, R. (1990). Concluding Essays on Applied Linguistics and Discourse Analysis. In R. Kaplan (Ed.), *Annual Review of Applied Linguistics*, Vol. 2.
- Krishnamurthy, R. (1996). Ethnic, racial and tribal: The language of racism? In Caldas Coulthard CR and Coulthard M (Eds.), *Texts and Practices: Readings in Critical Discourse Analysis*: 129-149. London: Routledge.
- Louw, E. (2001). *The Media and Cultural Production*. London: Sage.
- MacRitichie, V. & Seedat, M. (2008). Headlines and discourses in newspaper reports on traffic accidents. *South African Journal of Psychology*, 38(2), 337-354.
- Metila, R. (2013). A discourse analysis of news headlines: diverse framing for a hostage-taking event. *Asian Journal of Social Sciences & Humanities*, Vol.2, Number 2: 71-78.
- Taiwo, R. (2004). Speech as Headline in Nigerian Newspapers. In S. Awonusi & E. A. Babalola (Eds.), *The Domestication of English in Nigeria*: 323-335. Lagos: University of Lagos Press.
- Van Dijk, T. A. (1995). Discourse semantics as ideology. *Discourse & Society*, 6(2): 243-289.
- White, M. (2011). Cracking the code of press headlines: From difficulty to opportunity for the foreign language learner. *International Journal of English studies*, 11(1).